Gestural Recognition of Spanish Sign Language Vowels

Vanessa Alvear
vanessa.alvear@irsoluciones.com
Innovación Riojana de Soluciones IT S.L., and University of La Rioja
Logroño, La Rioja, SPAIN

ABSTRACT

Human gesture and action recognition (HAR) is a research topic that has been gaining attention thanks to the advances in cameras, technologies, and machine learning methods. This field develops techniques trying to establish interactions between humans and machines by means of gestures and actions. The field has many applications such as healthcare, prevention of dangerous activities, safety with smart surveillance, and sign language recognition, among others. The aim of the sign language recognition is to identify hand gestures used in sign language, thereby aiding the general public in understanding the communication efforts of individuals who are deaf or mute. As with the oral language, different sign languages exist in different countries. This paper presents our first approach to develop a gestural recognition of Spanish Sign Language. In particular, we have created a gesture recognition system of the vowels in Spanish Sign Language (Lengua de Signos Española - LSE).

CSCS CONCEPTS

• Human-centered computing → Gestural input.

KEYWORDS

Gesture Recognition, HAR, Spanish Sign Language, Computer Vision

ACM Reference Format:

1 INTRODUCTION

In the last decade, HAR tasks have been carrying out wide-ranging research works [5]. Furthermore, the recent spread of low-cost video camera systems, including depth cameras has improved the development of systems that require HAR in real-time. Some real-life applications are surveillance, home security, and healthcare, among others [5], which usually are studied in closed or controlled environments.

This work is included in a more general goal to provide human-machine interaction by recognizing actions in real time using cameras that capture images in open, or uncontrolled environments (e.g., a shopping mall, a university lobby, or a train station platform) allowing passers-by to pass through these spaces. This task presents challenges in technological, social, and legal aspects which need to be thoroughly analyzed. In this work, we present our first approach to develop a gestural recognition of Spanish Sign Language and we want to learn about it in order to apply it to the more general aim. It integrates a thesis project done as an Industrial Doctorate due to the collaboration between the Spanish company Innovación Riojana de Soluciones IT S.L. and the University of La Rioja.

2 DATA ANALYSIS

It is worth mentioning that sign language is not universal, and each country has its own sign language. Most works are focused on English Sign Language [6] and Spanish one is not so much studied.

In this work, we focus on five gestures corresponding to LSE vowels and create a recognition system in a controlled environment in two phases. In the first phase, we create a balanced dataset named \textit{vowelsLSE}. The dataset consists of 5 gestures of one person signing each vowel according to LSE and contains 3461 images. It consists of RGB images in JPG format with a size of 400 × 400 and has a white background to make them the same size. For more information, see the Table 1. The images were taken by a single person, and the sign was captured by a conventional webcam. In addition, it has images of both hands in different scenarios, with different positions of the hand relative to the body, and varying lighting (natural or artificial), etc.

In the second phase, we chose a technique from the literature review based on detection and classification models. To do so, we use the pipeline depicted in Figure 1, where we have the person doing the sign, then apply the hand detection model used from \textit{FastAI} [2], and \textit{Keras} [1], respectively.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Right Hand</th>
<th>Left Hand</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>332</td>
<td>335</td>
<td>667</td>
</tr>
<tr>
<td>E</td>
<td>377</td>
<td>393</td>
<td>770</td>
</tr>
<tr>
<td>I</td>
<td>333</td>
<td>335</td>
<td>668</td>
</tr>
<tr>
<td>O</td>
<td>340</td>
<td>338</td>
<td>678</td>
</tr>
<tr>
<td>U</td>
<td>334</td>
<td>344</td>
<td>678</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1716</strong></td>
<td><strong>1745</strong></td>
<td><strong>3461</strong></td>
</tr>
</tbody>
</table>

Table 1: \textit{vowelsLSE} dataset information.
Figure 1: Pipeline used for LSE vowel recognition. Example of the vowel U.

3 PRELIMINARY RESULTS

For that, we split data into 70%, 10%, and 20% for training, validation, and testing, respectively. The results of the evaluation are included in the Table 2. The best accuracy was 99.2% by ResNet50 model, followed by 96.5% by NN2. The dataset and the programs created are available on the website https://github.com/AlvearVanessa/vowelsLSE_signRecognition/tree/main.

Table 2: Results of evaluation.

<table>
<thead>
<tr>
<th>Model</th>
<th>Training Time</th>
<th>Accuracy test set (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN2</td>
<td>3 min 14 seg</td>
<td>96.5</td>
</tr>
<tr>
<td>ResNet50</td>
<td>2 min 41 seg</td>
<td>99.2</td>
</tr>
</tbody>
</table>

4 CONCLUSIONS

We have performed a proof of concept to create a dataset and a recognition system that allows us to become familiar with these methods. Going forward, we want to extend the study to uncontrolled environments through a contactless interface, which will also involve important issues such as those related to privacy. As well as, increase the human actions to be treated, adding variability to data, and the reliability of the people who are part of the dataset, addressing issues of race and gender. This project is designed to have a relevant impact both on the company where it is developed and on the research carried out by the university, without losing sight of the fact that this technology can ultimately have an impact on society.

5 ACKNOWLEDGMENTS

We would like to thank the company Innovación Riojana de Soluciones IT S.L., the University of La Rioja, and the Government of La Rioja, and mention that this work was partially supported by industrial doctorate, Comunidad Autónoma de La Rioja, Grant PID2020-115225RB-I00 funded by MCIN/AEI/10.13039/50110011033 and Proyecto Inicia 2023/01 funded by the Comunidad Autónoma de La Rioja.

REFERENCES